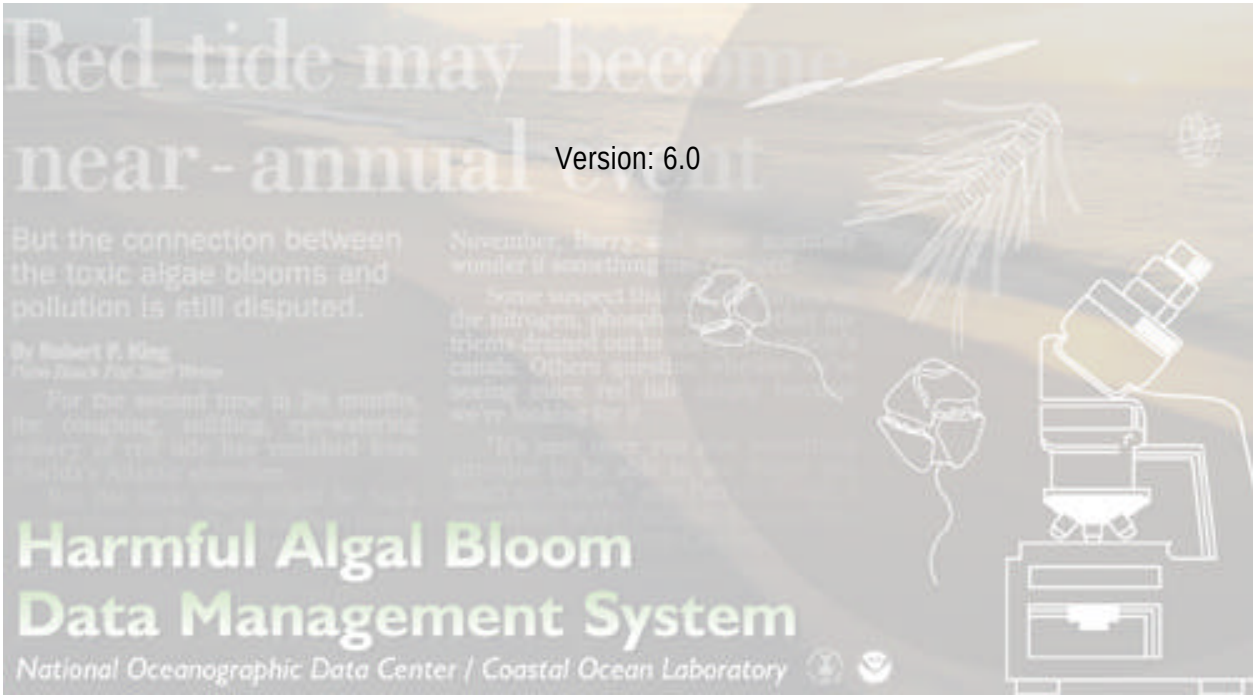
Harmful Algal Bloom Data Management System (HAB-DMS)

Exchange Format



Version: 6.0

HAB Team
National Oceanographic Data Center, NOAA
Silver Spring, Maryland
June 21, 2000

Table of Contents

I. Purpose of Document

Due to the spatial extent in which harmful algal events occur, it is important to understand the ecosystem dynamics relating to these events along coasts in which multiple states border. Although many states investigate harmful algal events, there is generally little collaboration towards the integration of these disparate data sources. Therefore, a data format has been developed at the NODC to facilitate the exchange and integration of data between multiple state and academic programs. In addition, software will be developed from this format for loading data into the Harmful Algal Bloom Data Management System (HAB-DMS). To provide timely access to data from the HAB-DMS, this exchange format should be used for data submission to the NODC. The purpose of this document is to provide background information regarding the structure of the database tables and to discuss the exchange format in which data should be submitted to the NODC. In addition, data retrieved from the HAB database will be provided in this format.

II. Database Description

Following is a description of the structure of the HAB database, with detailed information regarding each table. They have been included within this document in order to demonstrate the structure and fields by which the data are stored. A "record" in the exchange format is used to load each table, which represents a particular stage in the process of collecting and analyzing HAB measurements (Table 1). Since each data set is unique in the types of metadata which are associated with a measurement we have structured the database so that it is flexible in the amount of information that will be stored per data entry. We encourage data providers to submit as much information as possible, in the exchange format, so that the data are useful to other users. Most of the tables contain a remark field in which additional information about the dataset can be stored.

Table 1. A summary of each table in the HAB database and the types of data they contain.

Table	Data Types
EVENT	General information regarding a data submission (NODC accession)
CRUISE	Describes the sampling regime when associated with a cruise
COLLECTION	The where, when and how a sample was collected from the water column
ENVIRONMENT	Weather-related and general observations collected at a station

CHEM_ANAL	Further describes methods used in the chemical analysis of a sample and stores all chemical parameters
BIOLOGY	Further describes methods used in analyzing biological samples
BIO_DESC	Contains further descriptive and taxonomic information regarding a biological measurement, as well as all biological parameters

The following discussion describes the structure of the individual database tables.

EVENT Table

Data submitted to the NODC may be representative of various time and spatial scales. One submission is considered an accession and gets a unique number for archival purposes. The EVENT table summarizes some general information regarding the entire data set (the who, where and when for the entire accession).

Name	Format	Description
event_no	number	random number created by the database when data is loaded *(leave blank in format)*
project	number	project code: see project code table
cruise	number	cruise code: see cruise code table
investigator	number	investigator code (the person who submitted data): see investigator code table
source	char	institute code: see institute code table
region	char	custom-defined geographical region code: see region code table
site_type	char	fixed, or random site: see site_type code table
accession_no	char	accession number assigned by NODC (7 digits)
start_date	date/time	event start date (YYYYMMDD)
end_date	date/time	event end date (YYYYMMDD)

remark	memo	remark, other miscellaneous information entered during ingest (unlimited length text field)

CRUISE Table

This CRUISE table contains general information about the cruise regarding overall time period, researchers, platform, and project. A record will only be submitted if the sampling regime was associated with specific cruises.

Name	Format	Description
event_no	number	random number created by the database when data is loaded *(leave blank in format)*
cruise_id	char	cruise id generated by database or provided by investigator
start_date	date/time	cruise start date (YYYYMMDD)
end_date	date/time	cruse end date (YYYYMMDD)
chief_scientist	char	person in charge of cruise: use investigator code: see investigator code table
platform	char	platform (i.e., vessel/aircraft/vehicle name up to 30 characters in length)
project_no	number	project code: see project code table
institute	number	institute code: see institute code table

COLLECTION Table

Each row of the COLLECTION table contains information pertaining to where, when, and how a sample was collected, at the station level.

Name	Format	Description
event_no	number	random number created by the database when data is loaded *(leave blank in format)*
collection_no	number	random number created by the database when data is loaded *(leave blank in format)*

station_no	number	station code: see station code table
start_date	date/time	start date/time of collection
end_date	date/time	end date/time of collection
tow_distance	number	distance of tow/haul in meters
latitude	char	latitude of collection location +/-DD.DDDDD (- is south)
longitude	char	longitude of collection location +/-DDD.DDDDD (- is west)
ll_datum	char	ll_datum code: see ll_datum table
upper_depth	number	upper depth of sample collection
lower_depth	number	lower depth sample collection
total_depth	number	total station depth
z_unit	number	unit code: see unit code table
gear_no	number	gear code: see gear code table
layer	char	layer code: see layer code table

ENVIRONMENT Table

The ENVRIONMENT table contains weather-related information, such as wind speed and direction, precipitation, water color, etc. There will be one record for every weather-related parameter that was measured during collection.

Name	Format	Description
event_no	number	random number created by the database when data is loaded *(leave blank in format)*
collection_no	number	collection code: see collection code table
parameter_no	number	parameter code: see parameter code table
value	number	parameter value
unit_no	number	unit code: see unit code table
remarks_no	number	remark code: see remark code table

CHEM_ANAL Table

This table contains information about the chemical analysis of a sample within the laboratory.

Name	Format	Description
event_no	number	random number created by the database when data is loaded *(leave blank in format)*
collection_no	number	collection code: see collection code table
sample_no	number	random number created by the database when data is loaded *(leave blank in format)*
sample_type	char	sample type code: see sample_type code table
orig_sample_code	char	originator's sample code/number (up to 20 characters)
parameter_no	number	parameter code: see parameter code table
value	number	parameter value
unit_no	number	unit code: see unit code
qualifier	char	parameter detection limit qualifier code: see detection limit qualifier code table
lab	number	institute code: see institute code table
analyst	number	investigator code: see investigator code table
method_no	number	method code: see method code table
remark_no	number	remark code: see remark code table

BIOLOGY Table

This table contains information regarding biological samples that are further analyzed within a laboratory. It describes how the sample was analyzed, who analyzed it, etc.

Name	Format	Description

event_no	number	random number created by the database when data is loaded *(leave blank in format)*
collection_no	number	collection code: see collection code table
sample_no	number	random number created by the database when data is loaded *(leave blank in format)*
sample_type	char	sample_type code: see sample_type code table
orig_sample_code	char	originator's sample code or number (up to 20 characters)
parameter_no	number	parameter code: see parameter code table
tsn	number	taxonomy code: see taxonomy code table
analyst	number	investigator code: see investigator code table
method_no	number	method code: see method code table
bio_group_no	number	bio_group code: see bio_group code table
remark_no	number	remark code: see remark code table

BIO_DESC Table

The BIO_DESC table breaks the biological parameters down by sex, life_stage, etc. Parameter values, such as biomass, counts etc. are stored within this table, along with the units that are associated with that measurement. A remark field is also available if further description of the organism is needed.

Name	Format	Description
event_no	number	random number created by the database when data is loaded *(leave blank in format)*
collection_no	number	collection code: see collection code table
sample_no	number	random number created by the database when data is loaded *(leave blank in format)*
parameter_no	number	parameter code: see parameter code table
tsn	number	taxonomy code: see taxonomy code table

desc_no	number	biological description number, random number created by the database when data is loaded *(leave blank in format)*
modifier	char	taxonomic modifier (e.g., sp, spp) text (up to 3 characters)
life_stage_no	number	life_stage code: see life stage code table
sex_no	number	sex code: see sex code table
value	number	parameter value
unit_no	number	unit code: see unit code table
remark	memo	remark, other miscellaneous information entered during ingest (unlimited length text field)

III. HAB-DMS Exchange Format

The format that has been developed by the NODC for data entry into the HAB database has been designed to provide information regarding the collection and laboratory phases of data acquisition, associated with each chemical and biological parameter. This format was developed by the NODC to provide enough flexibility for data providers, while still maintaining some consistency in formatting. The HAB-DMS exchange format was designed to work well with spreadsheet or columnar data. Codes for data elements should be consistent with the NODC HAB-DMS codes, which were designed to match CIMS 2000 codes wherever possible. If a desired entry is not found in a HAB-DMS code table, place the full text within the format and one will be assigned by the NODC. Additional information can be added to the parameter records, as they pertain to the table structure (e.g., if additional information is available regarding the lifestage, sex, tow_distance, etc of a biological measurement, this information should be stored at the data level, with the appropriate header to describe the field). We have used a colon to separate field names with the data, to avoid any confusion regarding the beginning of a record. Each new record should be entered on a new line, where multiple entries of the same record type may follow on consecutive lines (e.g., if more than one Principal Investigator was responsible for the project, multiple entries (rows) should be entered to completely describe the data set). Field names should be spelled out (no=number, #=number) to avoid any discrepancies prior to loading. Codes should be consistent with the HAB-DMS data dictionary. We prefer that data submitted to NODC is pipe or comma delimited, however, any delimiter can be used as long as it is described within the *file info* record.

<file info>
original file: *name of the original file*
current file: *name of the file actually converted after titles, etc. are cut out*
converter: *program used to convert into this format*
delimiter: *the character used to separate columns of data*

<event info>
event number: *random number created by the database when data is loaded*
project: *project code: see project code table*
cruise: *cruise code: see cruise code table*
investigator: *investigator code (the person who submitted data): see investigator code table*
source: *institute code: see institute code table*
region: *custom-defined geographical region code: see region code table*
site type: *fixed, or random site: see site_type code table*
accession number: *accession number assigned by NODC (7 digits)*
start date: *event start date (YYYYMMDD)*
end date: *event end date (YYYYMMDD)*

remark: *remark, other miscellaneous information entered during ingest: text field*

<**environment info**>
column number(start with 0)|parameter code|units|remarks

<station info> *This information should be provided one time, and when station locations change for a program.
station id|latitude|longitude|ll_datum

<chemical parameter info>
column number(start with 0)|parameter code|layer|units|qualifier|lab|investigator|gear|method|remark

<biological parameter info>
column number(start with 0)|parameter code|layer|units|tsn|bio_group|investigator|gear|method|remark

<header info>
delimited version of the header column of the spreadsheet

<data>
delimited version of the spreadsheet contents

IV. Sample Chemical Data Set

A sample of DE-DNREC Pfiesteria monitoring data in MS Excel format was used to demonstrate the use of the HAB-DMS exchange format. The original has more parameters extending to the right, and more stations below. The following data set is a subset which represents chemical data from the DE-DNREC Pfiesteria monitoring program.

Station	ESS Sample Number	Date Sampled	Time Sampled	Total Depth	TSS	Chl-a	TP
I-1	98013400	5/5/98	10:27		47.0	5.0	0.084
I-1		5/19/98	9:25	10.0	39.0	8.0	0.14
I-1	98018300	6/9/98	11:00	4.5	33.0	3.0	0.152
I-1	98021660	6/22/98	12:00	3.5	43.0	3.0	0.114
I-1	98025130	7/6/98	9:53	5.5	74.0	3.0	0.204
I-1	98029630	7/22/98	11:00	5.5	40.0	8.0	0.446
I-1	98033300	8/5/98	15:50	3.5	33.0	5.0	0.076
I-1	98036070	8/18/98	10:58	6.0	135.0	11.0	0.098
I-1	98038570	8/31/98	11:15	5.5	147.0	16.0	0.456
I-1	98041640	9/14/98	12:35	5.0	114.0	8.0	0.078

I-1	98044690	9/30/98	10:40	5.0	120.0	5.0	
I-1	98049300	10/21/98	11:00	5.8	31.0	8.0	
IP-1	98013490	5/5/98	13:06	6.0	38.0	24.0	0.121

Sample of Format: Using the same spreadsheet data above.

<file info>
original file: rt990212.xls
current file: /disk4/hab/data/testdata.csv
converter: de_p1.pro
delimiter: pipe
<event info>
event number: H0000001
project: DE-DNREC Pfiesteria Monitoring
cruise:
investigator: Edythe Humphries
source: DE-DNREC
region: Chesapeake
site type: fixed
accession number: 0000001
start date:19980505
end date: 19981021
remark: sample dataset - testing only
<chemical parameter info>
4|TSS|S|mg/l||DNREC-ELS-ASB|Kathy Knowles||TSS|
5|CHLA|S|ug/l||DNREC-ELS-ASB|Dave Saveikis||PHEO1|
6|TP|B|mg/l||DNREC-ELS-ASB|Kathy Knowles||PHOS1|
<header info>
station id|sample no|date|time|total depth|TSS|CHLA|TP
<data>
I-1|98013400|5/5/98|10:27||47.0|5.0|0.084
I-1||5/19/98|9:25|10.0|39.0|8.0|0.14
I-1|98018300|6/9/98|11:00|4.5|33.0|3.0|0.152
I-1|98021660|6/22/98|12:00|3.5|43.0|3.0|0.114
I-1|98025130|7/6/98|9:53|5.5|74.0|3.0|0.204
I-1|98029630|7/22/98|11:00|5.5|40.0|8.0|0.446
I-1|98033300|8/5/98|15:50|3.5|33.0|5.0|0.076
I-1|98036070|8/18/98|10:58|6.0|135.0|11.0|0.098
I-1|98038570|8/31/98|11:15|5.5|147.0|16.0|0.456
I-1|98041640|9/14/98|12:35|5.0|114.0|8.0|0.078

I-1|98044690|9/30/98|10:40|5.0|120.0|5.0|
I-1|98049300|10/21/98|11:00|5.8|31.0|8.0|
IP-1 |98013490|5/5/98|13:06|6.0|38.0|24.0|0.121

V. Sample Biological Data Set (without taxonomic information)

The following data set represents a subset of data from DE-DNREC's Pfiesteria monitoring program. It represents data which contains biological information, however, the biological data is not analyzed to the taxonomic level (does not contain a tsn). Therefore, PLO is identified as a bio_group in this format.

Station	ESS Sample Number	Date Sampled	Time Sampled	Total Depth	PLO (#/ml)	Code	Analyst
I-1	98013400	5/5/98	10:27			None	EH
I-1		5/19/98	9:25	10.0		None	EH
I-1	98018300	6/9/98	11:00	4.5	10		GMM
I-1	98021660	6/22/98	12:00	3.5		None	GMM
I-1	98025130	7/6/98	9:53	5.5	10		GMM
I-1	98029630	7/22/98	11:00	5.5		None	GMM
I-1	98033300	8/5/98	15:50	3.5		None	GMM
I-1	98036070	8/18/98	10:58	6.0		None	GMM
I-1	98038570	8/31/98	11:15	5.5		None	GMM
I-1	98041640	9/14/98	12:35	5.0		None	GMM
I-1	98044690	9/30/98	10:40	5.0		None	GMM

<file info>
original file: rt990212.xls
current file: /disk4/hab/data/testdata.csv
converter: de_p2.pro
delimiter: pipe
<event info>
event number: H0000002
project: DE-DNREC Pfiesteria Monitoring
cruise:
investigator: Edythe Humphries
source: DE-DNREC
region: Chesapeake
site type: fixed
accession number: 0000002
start date:19980505
end date: 19981021
remark: sample dataset - testing only
<biological parameter info>
4|COUNT|S|number/ml||PLO|Edythe Humphries|Whole Water Column Sampler|PLO1|a value of 0 indicates that no Pfiesteria like organisms were observed in 0.1 ml subsample of Lugol's preserved 125 ml sample

<header info>
station id|sample number|date|time|total depth|PLO
<data>
I-1|98013400|5/5/98|10:27|0
I-1||5/19/98|9:25|0
I-1|98018300|6/9/98|11:00|10
I-1|98021660|6/22/98|12:00|0
I-1|98025130|7/6/98|9:53|10
I-1|98029630|7/22/98|11:00|0
I-1|98033300|8/5/98|15:50|0
I-1|98036070|8/18/98|10:58|0
I-1|98038570|8/31/98|11:15|0
I-1|98041640|9/14/98|12:35|0
I-1|98044690|9/30/98|10:40|0

Sample Biological Data Set (with taxonomic information)

The following example represents sample data collected during the January 1999 ECOHAB-Florida cruise and was provided by Dr. Gabe Vargo for use in the development of the HAB database at NODC. It represents a biological data set which also contains taxonomic information.

DATE	TIME	LAT	LONG	Station	Depth (m)	G. breve (cells/l)
1/13/1999	4:17	27.465	-82.966	3	0	0
1/13/1999	4:17	27.465	-82.966	3	13	0
1/13/1999	2:54	27.389	-83.134	5	0	0
1/13/1999	2:54	27.389	-83.134	5	25	0
1/13/1999	1:26	27.314	-83.301	7	0	0
1/13/1999	1:26	27.314	-83.301	7	30	0
1/13/1999	23:59	27.238	-83.468	9	0	0
1/13/1999	23:10	27.2	-83.552	10	0	0
1/13/1999	23:10	27.2	-83.552	10	40	0
1/13/1999	15:57	26.472	-84.392	11	0	0

<file info>
original file: eh0199gbreveconcentration.txt
current file: /disk4/hab/data/tabgbreve.txt
converter: gbreve_sample.pro
delimiter: comma
<event info>
event number: H0000003
project:ECOHAB-Florida
cruise:
investigator: Dr. Gabriel Vargo
source: Department of Marine Science, University of South Florida
region: Gulf of Mexico
site type: random
accession number: 0000003
start date:19990113
end date: 19990113
remark: sample dataset - testing only
<biological parameter info>
column no(start with 0),parameter code,units,tsn,bio group,investigator,gear,method,remark
4,COUNT,D,number/liter,10157,,Dr. Gabriel Vargo,rosette sampler, method for g. breve,
<header info>

date,time,latitude,longitude,station,depth,G. breve
<data>
1/13/1999,4:17,27.465,-82.966,3,0,0
1/13/1999,4:17,27.465,-82.966,3,13,0
1/13/1999,2:54,27.389,-83.134,5,0,0
1/13/1999,2:54,27.389,-83.134,5,25,0
1/13/1999,1:26,27.314,-83.301,7,0,0
1/13/1999,1:26,27.314,-83.301,7,30,0
1/13/1999,23:59,27.238,-83.468,9,0,0
1/13/1999,23:10,27.2,-83.552 10,0,0
1/13/1999,23:10,27.2,-83.552 10,40,0
1/13/1999,15:57,26.472,-84.392,11,0,0